CLAIMS

1)Method of discretization / grouping of a source attribute or a source attributes group of a database containing a population of individuals with the object in particular of predicting modalities of a given target attribute, said method comprising the following steps of:

a)Partition of said modalities of said source attribute or said attribute group into elementary regions,

b)Evaluation of a merge criterion for each pair of elementary regions,

c)Search, among the set of pairs of elementary regions that can be merged, for the pair of elementary regions for which the merge criterion would be optimized,

e)Stopping of the method if there are no elementary regions whose merge would have the consequence of improving said merge criterion,

f)otherwise merge and reiteration of steps b) to e),

characterized in that it comprises in addition a step d) between steps c) and e) that skips directly to step f) as long as the value of a valuation variable of the merge under consideration, said valuation variable characterizing the behavior of said merge criterion, is not comprised in a predetermined zone of atypical values.

2)Method of discretization / grouping of a source attribute or source attributes group according to claim 1, characterized in that said predetermined zone of atypical values is such that for a target attribute independent of said source attribute or said source attributes group, the value of said valuation variable of the merge under consideration is not comprised in said zone with a predetermined probability p.

3)Method of discretization of a source attribute of a database containing a population of individuals with the object in particular of predicting modalities of a given target attribute, said method comprising the following steps of:

a)Partition of said modalities of the source attribute into adjacent two-by-two elementary intervals.

b)Evaluation for each pair of adjacent elementary intervals of said set of the value of χ^2 of the contingence table after a possible merge of said pair,

c)Search, among the set of pairs of elementary intervals that can be merged, for the pair of elementary intervals whose merge would maximize the value of χ^2,

e)Stopping of the method if there are no elementary intervals that make it possible to reduce the probability of independence,

f)otherwise merge and reiteration of steps b) to e),

characterized in that it comprises in addition a step d) between steps c) and e) that skips directly to step f) as long as the value $\Delta\chi^2$ of the variation of the value of χ^2 before and after merge is, in absolute value, less than a predetermined threshold value $Max\Delta\chi^2$.

4)Discretization method according to claim 3, characterized in that said predetermined threshold value $Max\Delta\chi^2$ is such that for a target attribute independent of the source attribute the value $\Delta\chi^2$ of the variation of the value of χ^2 before and after merge is always less than said value $Max\Delta\chi^2$ with a predetermined probability p.

5)Discretization method according to claim 4), characterized in that said predetermined threshold value $Max\Delta\chi^2$ is equal to the function of χ^2 of degree of freedom equal to the number J of modalities of the target attribute minus one for a probability p to the power 1/N where N is the size of the sample of the part of the database to which said discretization method is applied:

$$Max\Delta\chi^2 = Inv\chi^2_{J-1}\left(p^{1/N}\right)$$

where $Inv\chi^2$ is the function that gives the value of χ^2 as a function of a given probability p.

6)Method of discretization of a source attribute according to one of claims 3 to 5, characterized in that it comprises a step of verification that the effective of a source attribute for modalities in a given interval for each target attribute is greater than a predetermined value, and if such is not the case, to implement the merge of said interval with an adjacent interval.

7)Method of grouping of a source attribute of a database containing a population of individuals with the object in particular of predicting

modalities of a given target attribute, said method comprising the following steps of:

a)Partition of said modalities of the source attribute into a plurality of groups,

b)Evaluation for each pair of groups of said set of the value of χ^2 of the contingence table after a possible merge of said pair,

c)Search, among the set of pairs of groups that can be merged, for the pair of groups whose merge would maximize the value of χ^2,

e)Stopping of the method if there are no merges of groups that make it possible to reduce the probability of independence,

f)otherwise merge and reiteration of steps b) to e),

characterized in that it comprises in addition a step d) between steps c) and e) that skips directly to step f) as long as the value $\Delta\chi^2$ of the variation of the value of χ^2 before and after merge is, in absolute value, less than a predetermined threshold value $\text{Max}\Delta\chi^2$.

8)Grouping method according to claim 7, characterized in that said predetermined threshold value $\text{Max}\Delta\chi^2$ is such that for a target attribute independent of the source attribute the value $\Delta\chi^2$ of the variation of the value of χ^2 before and after merge is always less than said value $\text{Max}\Delta\chi^2$ with a predetermined probability p.

9)Grouping method according to claim 7, characterized in that to establish the predetermined threshold value $\text{Max}\Delta\chi^2$, it consists in using a previously calculated table of values of mean and standard deviation as a function of the number of modalities of the source attribute and of the number of modalities of the target attributes to determine by linear interpolation from said table of values the mean and standard deviation of $\text{Max}\Delta\chi^2$ corresponding to the attributes to be grouped, and then to determine, by using the inverse normal law, the corresponding predetermined threshold value $\text{Max}\Delta\chi^2$ which will not be with a probability p.

10)Grouping method according to claim 9, characterized in that for two target modalities, the mean of $\text{Max}\Delta\chi^2$ is asymptotically proportional to $2I/\pi$, where I is the number of source modalities.

11)Grouping method according to claim 10), characterized in that for two source modalities, the law of $Max\Delta\chi^2$ is the law of χ^2 with J-1 degrees of freedom, J being the number of target modalities.

12)Method of grouping of a source attribute according to one of the preceding claims 7 to 11, characterized in that it comprises a preliminary step of verification that the effective of a source attribute for modalities in a given group for each target attribute is greater than a predetermined value, and if such is not the case, to implement a merge of said group with a specific group, said merged group then forming again said specific group.

13)Method of discretization in dimension k of a group of k continuous source attributes of a database containing a population of individuals, with the object in particular of predicting the modalities of a given target attribute, said method comprising the following steps of:

a)Partition of said modalities of the group of k source attributes into elementary regions of dimension k,

b)Evaluation for each pair of adjacent elementary regions of the value of χ^2 of the contingence table after a possible merge of said pair,

c)Search, among the set of pairs of regions that can be merged, for the pair of regions whose merge would maximize the value of χ^2,

e)Stopping of the method if there is no set of intervals that make it possible to reduce the probability of independence,

f)otherwise merge and reiteration of steps b) to e),

characterized in that it comprises in addition a step d) between steps c) and e) that skips directly to step f) as long as the value $\Delta\chi^2$ of the variation of the value of χ^2 before and after merge is, in absolute value, less than a predetermined threshold value $Max\Delta\chi^2$.

14)Method of grouping in dimension k of a group of k discrete source attributes of a database containing a population of individuals, with the object in particular of predicting the modalities of a given target attribute, said method comprising the following steps of:

a)Partition of said modalities of the group of k source attributes into a plurality of groups,

b)Evaluation for each pair of groups of the value of χ^2 of the contingence table after a possible merge of said pair,

c)Search, among the set of pairs of groups that can be merged, for the pair of groups whose merge would maximize the value of χ^2,

e)Stopping of the method if there are no merges of groups that make it possible to reduce the probability of independence,

f)otherwise reiteration of steps b) to e),

characterized in that it comprises in addition a step d) between steps c) and e) that skips directly to step f) as long as the value $\Delta\chi^2$ of the variation of the value of χ^2 before and after merge is, in absolute value, less than a predetermined threshold value $Max\Delta\chi^2$.